1/5/1

DIALOG(R)File 351:Derwent WPI

010338739 **Image available**
WPI Acc No: 1995-240827/199531
XRPX Acc No: N95-187756

Voice recognition system appts. for speech signal processing - extracts features from input speech frame to supply word decoder in central processing station

Patent Assignee: QUALCOMM INC (QUAL-N)
Inventor: CHANG C; JACOBS P E
Number of Countries: 064 Number of Patents: 019
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
WO 9517746	A1	19950629	WO 94US14803	A	19941220	199531	B
AU 9513753	A	19950710	AU 9513753	A	19941220	199543	
ZA 9408426	A	19950927	ZA 948426	A	19941026	199544	
EP 736211	A1	19961009	WO 94US14803	A	19941220	199645	
			EP 95904956	A	19941220		
FI 9602572	A	19960820	WO 94US14803	A	19941220	199646	
			FI 962572	A	19960620		
BR 9408413	A	19970805	BR 948413	A	19941220	199738	
			WO 94US14803	A	19941220		
JP 9507105	W	19970715	WO 94US14803	A	19941220	199738	
			JP 95517605	A	19941220		
KR 97700353	A	19970108	WO 94US14803	A	19941220	199801	
			KR 96703304	A	19960621		
CN 1138386	A	19961218	CN 94194566	A	19941220	199806	
TW 318239	A	19971021	TW 94110578	A	19941115	199808	
AU 692820	B	19980618	AU 9513753	A	19941220	199835	
IL 112057	A	19981126	IL 112057	A	19941219	199912	
US 5956683	A	19990921	US 93173247	A	19931222	199945	
			US 95534080	A	19950921		
			US 96627333	A	19960404		
KR 316077	B	20020228	WO 94US14803	A	19941220	200260	
			KR 96703304	A	19960621		
US 6594628	B1	20030715	US 95534080	A	19950921	200348	N
			US 96627333	A	19960404		
			US 97832581	A	19970402		
MX 208881	B	20020712	MX 95184	A	19950102	200366	
EP 1381029	A1	20040114	EP 95904956	A	19941220	200410	
			EP 200321806	A	19941220		
EP 736211	B1	20040303	WO 94US14803	A	19941220	200417	
			EP 95904956	A	19941220		
			EP 200321806	A	19941220		
DE 69433593	E	20040408	DE 633593	A	19941220	200425	
			WO 94US14803	A	19941220		
			EP 95904956	A	19941220		

Priority Applications (No Type Date): US 93173247 A 19931222; US 95534080 A 19950921; US 96627333 A 19960404; US 97832581 A 19970402
Cited Patents: EP 108354; EP 177405; EP 534410
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
WO 9517746	A1	E	18	G10L-005/06	

Designated States (National): AM AT AU BB BG BR BY CA CH CN CZ DE DK EE ES FI GB GE HU JP KE KG KP KR KZ LK LR LT LU LV MD MG MN MW NL NO NZ PL PT RO RU SD SE SI SK TJ TT UA UZ VN
Designated States (Regional): AT BE CH DE DK ES FR GB GR IE IT KE LU MC MW NL OA PT SD SE SZ

AU 9513753	A		G10L-005/06	Based on patent WO 9517746
ZA 9408426	A	20	G10L-000/00	

EP 736211	A1 E	18	G10L-005/06	Based on patent WO 9517746
Designated States (Regional): AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE				
FI 9602572	A		G10L-000/00	
BR 9408413	A		G10L-005/06	Based on patent WO 9517746
JP 9507105	W	20	G10L-003/00	Based on patent WO 9517746
KR 97700353	A		G10L-005/06	Based on patent WO 9517746
CN 1138386	A		G10L-005/06	
TW 318239	A		G10L-007/08	
AU 692820	B		G10L-005/06	Previous Publ. patent AU 9513753 Based on patent WO 9517746
IL 112057	A		G10L-005/06	
US 5956683	A		G10L-003/00	Cont of application US 93173247 Cont of application US 95534080
KR 316077	B		G10L-015/00	Previous Publ. patent KR 97700353 Based on patent WO 9517746
US 6594628	B1		G10L-015/00	Cont of application US 95534080 Cont of application US 96627333
MX 208881	B		G10L-003/00	
EP 1381029	A1 E		G10L-015/28	Div ex application EP 95904956 Div ex patent EP 736211
Designated States (Regional): AT BE CH DE DK ES FR GB GR IE IT LI LT LU MC NL PT SE				
EP 736211	B1 E		G10L-015/02	Related to application EP 200321806 Related to patent EP 1381029 Based on patent WO 9517746
Designated States (Regional): AT BE CH DE DK ES FR GB GR IE IT LI LT LU MC NL PT SE SI				
DE 69433593	E		G10L-015/02	Based on patent EP 736211 Based on patent WO 9517746

Abstract (Basic): WO 9517746 A

The feature extraction apparatus (22) is located at a remote station (40) for receiving a frame of speech samples and extracting a set of speech features in accordance with a predetermined feature extraction format to provide the set of speech features.

A word decoder (48) located at a central processing station receives the set of speech features and determines a syntax in accordance with a predetermined decoding format. The set of features are linear predictive coding parameters. A local word detector may be collocated in the remote station for determining a syntax in accordance with a predetermined small vocabulary decoding format.

ADVANTAGE - Improved system performance is given by appropriately separating components of feature extraction and word decoding.

Dwg.2/5

Title Terms: VOICE; RECOGNISE; SYSTEM; APPARATUS; SPEECH; SIGNAL; PROCESS; EXTRACT; FEATURE; INPUT; SPEECH; FRAME; SUPPLY; WORD; DECODE; CENTRAL; PROCESS; STATION

Derwent Class: P86; W01; W04

International Patent Class (Main): G10L-000/00; G10L-003/00; G10L-005/06; G10L-007/08; G10L-015/00; G10L-015/02; G10L-015/28

International Patent Class (Additional): G10L-005/00; G10L-005/02; G10L-009/06; G10L-015/26

File Segment: EPI; EngPI

?

JPA09-507105 which corresponds to USP5,956,683 and USP6,594,628

(19)日本国特許庁（ＪＰ）　(12)**公表特許公報**（Ａ）　(11)特許出願公表番号

特表平9－507105

(43)公表日　平成９年(1997)７月15日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	
Ｇ１０Ｌ　3/00	５５１	9379－5H	Ｇ１０Ｌ　3/00	５５１Ａ

審査請求　未請求　予備審査請求　有　（全 20 頁）

(21)出願番号　特願平7－517605
(86)(22)出願日　平成６年(1994)12月20日
(85)翻訳文提出日　平成８年(1996)６月24日
(86)国際出願番号　ＰＣＴ／ＵＳ９４／１４８０３
(87)国際公開番号　ＷＯ９５／１７７４６
(87)国際公開日　平成７年(1995)６月29日
(31)優先権主張番号　１７３，２４７
(32)優先日　1993年12月22日
(33)優先権主張国　米国（ＵＳ）

(71)出願人　クゥアルコム・インコーポレーテッド
アメリカ合衆国、カリフォルニア州 92121、サン・ディエゴ、ラスク・プールバード　6455
(72)発明者　ジェイコブス、ポール・イー
アメリカ合衆国、カリフォルニア州 92037、ラ・ジョラ、ラ・ジョラ・ショアズ・レーン　9075
(72)発明者　チャン、シエンチュン
アメリカ合衆国、カリフォルニア州 92131、サン・ディエゴ、サイプレス・テラス　11456
(74)代理人　弁理士　鈴江　武彦　（外４名）

最終頁に続く

(54)【発明の名称】　分散音声認識システム

(57)【要約】
特徴抽出装置（２２）を有する音声認識システムがリモート局（４０）に設けられる。特徴抽出装置（２２）は入力音声フレームから特徴を抽出して抽出された特徴を中央処理局（４２）に供給する。中央処理局（４２）では、前記特徴がワード復号器（４８）に供給されて入力音声フレームのシンタックスが決定される。

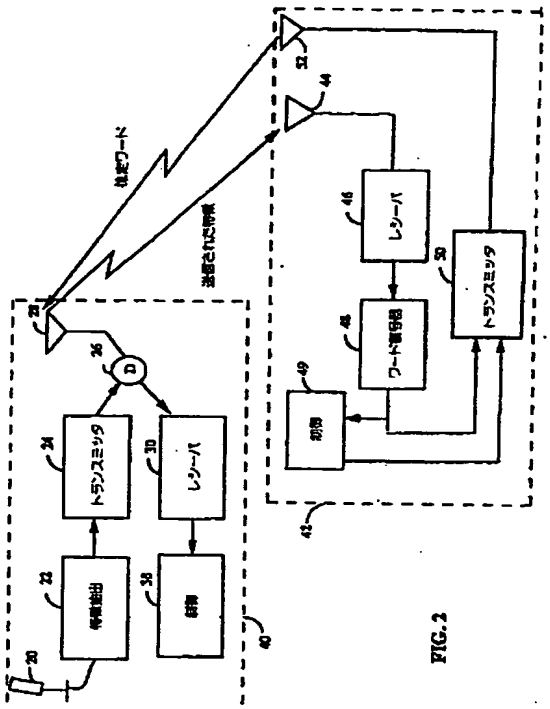

FIG. 2

【特許請求の範囲】

1．音声認識システムであって、

リモート局に設けられ、音声サンプルのフレームを受信して、所定の特徴抽出フォーマットに従って前記音声サンプルのフレームから、一対の音声特徴を抽出して、前記一対の音声特徴を提供するための特徴抽出手段と、

中央処理局に設けられ、前記一対の音声特徴を受信して、所定の復号フォーマットに基づいてシンタックスを決定するワード復号器と、

を具備することを特徴とする音声認識システム。

2．前記一対の音声特徴が線形予測符号化パラメータであることを特徴とする請求の範囲第1項に記載の音声認識システム。

3．前記リモート局に配置され、所定の小さい語彙の復元フォーマットに従ってシンタックスを決定するためのローカルワード検出器をさらに具備することを特徴とする請求の範囲第1項に記載の音声認識システム。

【発明の詳細な説明】

分散音声認識システム

発明の背景

１．発明の分野

本発明は音声信号処理に関する。特に、本発明は標準音声認識システムの分散実行を実現する新規な方法及び装置に関する。

２．関連技術の説明

音声認識は、ユーザまたはユーザ発生コマンドを認識、かつ、機械とのヒューマンインターフェースを達成するために、シュミレートされた知性を有する機械を提供する最も重要な技術である。また、ヒューマン音声理解に対する中心技術である。音響音声信号からの言語メッセージを復元する技術を使用するシステムは、音声認識装置（ＶＲ）と呼ばれている。音声認識装置は、到来する生の音声から、ＶＲに必要な一連の情報含有特徴（ベクトル）を抽出する音響プロセッサと、入力された音声に対する一連の言語ワードなどの、意味のある所望の出力フォーマットを得るために、前記一連の特徴（ベクトル）を復元するワード復号器とからなる。システムのパフォーマンスを増大させるために、システムに有効なパラメータを備えさせるトレーニングが必要である。すなわち、システムは最適に機能するようになるまで学習する必要がある。

音響プロセッサは音声認識装置におけるフロントエンド音声解析サブシステムを代表する。このシステムは入力音声信号に応答して時変音声信号を特徴付けるために、最適な表現を提供する。背景雑音、チャネルひずみ、話者特性や話し方などの無関係な情報は棄却される。効率のよい音響特徴は音声認識装置により高い音響識別能力を与える。最も有用な特性は短時間スペクトルエンベロープである。短時間スペクトルエンベロープを特徴付ける２つの最もよく用いられるスペクトル解析方法は、線形予測符号化（ＬＰＣ）モデルとフィルタバンクに基づくスペクトル解析モデルである。しかしながら、(Rabiner, L.R.及びSchafer, R.W.著、音声信号のディジタル処理、Prentice Hall,1978)に示されるように、ＬＰＣは音声軌跡(tract)スペクトルエンベロープに対するよい近似を提供するだ

けでなく、すべてのディジタル実行においてフィルタバンクモデルよりも計算上より安価である。経験によれば、ＬＰＣに基づいた音声認識装置のパフォーマンスは、フィルタバンクに基づく認識装置と同等かあるいはそれ以上である（Rabiner,L.R.及びB.H.著、音声認識の基本、Prentice Hall,1993）。

図１に示す、ＬＰＣに基づく音響プロセッサにおいて、入力音声がマイクロホン（図示せず）に供給されてアナログ電気信号に変換される。この電気信号はその後、（図示せぬ）Ａ／Ｄ変換器によってディジタル化される。このディジタル化された音声信号は、そのスペクトルを平らにして次の信号処理における有限プレシジョン効果(finite precision effects)を受けないようにすべく、プレエンファシスフィルタ２を通過される。プレエンファシスフィルタリングされた音声は区分要素(segmentation element)４に供給されて一時的に重複、または、重複しないブロックに区分、あるいはブロック化される。音声フレームデータは窓要素(windowing element)６に供給されてフレーム化されたＤＣ成分が除去されるとともに、フレーム境界における不連続によるブロッキング効果を低減するために、各フレームに関してディジタル窓処理が行われる。ＬＰＣ解析において最もよく使用される窓関数はハミング窓ｗ（ｎ）であり、以下のごとく定義される。

$$w(n)=0.54-0.46\cdot\cos\left(\frac{2\pi n}{N-1}\right),0\le n\le N-1 \qquad (1)$$

窓処理された音声はＬＰＣ解析要素８に供給される。ＬＰＣ解析要素８では、自己相関関数が窓処理されたサンプルに基づいて計算され、対応するＬＰＣパラメータが自己相関関数から直接得られる。

概して、ワード復号器は音響プロセッサによって生成された音響特徴シーケンスを話者の元ワード列に変換する。これは２つの工程、すなわち、音響パターンマッチングと言語モデリングにより達成される。言語モデリングは分離されたワード認識への応用では用いられない。ＬＰＣ解析要素８からのＬＰＣパラメータは音素、音節、ワードなどの可能な音響パターンを検出して分類する。候補パターンが言語モデリング要素１２に供給されて、ワードのどのシーケンスが文法的によく形成されかつ意味をもつかを決定する、シンタクス上の拘束(syntactic c

onstraints)の規則をモデル化する。音響情報自身があいまいである場合は、シンタクス情報は貴重な指針となる。言語モデリングに基づいて、ＶＲは逐次音響特徴マッチング結果を解釈して推定ワード列を提供する。

ワード復号器における音響パターンマッチングと言語モデリングは、話者の音声学上の及び音響音声学上の変化を記載するために、確定または確率的な数学モデルを必要とする。音声認識システムのパフォーマンスはこれらの２つのモデリングの品質に直接関連する。音響パターンマッチングのための種々のクラスのモデルのうち、テンプレートに基づくダイナミックタイムワーピング（ＤＴＷ）と確率的隠れマルコフモデリング（ＨＭＭ）とは最もよく用いられている２つの方法である。しかしながら、ＤＴＷに基づく方法はＨＭＭに基づく方法の特別な場合であるとみなすことができ、パラメータを用いた二重に確率的な(parametric doubly stochastic model)モデルである。ＨＭＭシステムは現在最も成功した音声認識アルゴリズムである。ＨＭＭにおける一重(doubly)特性は音響のみならず音声信号に関連した一時的変化を吸収するのにより大きな柔軟性を有している。これは改善された認識の正確さにつながる。言語モデルにおいて、ｋグラム言語モデルと呼ばれる確率モデルが実際的な大きな語彙の音声認識システムに適用された。この確率モデルはF.Jelink著、実験的離散デクテーション認識装置、Proc.IEEE,vol.73,pp.1616-1624に詳細に述べられている。一方、小さいな語彙の場合は、確定的文法が、航空及び予約及び情報システムへの応用において、有限状態ネットワーク（ＦＳＮ）として確立されている(Rabiner,L.R.及びLevinson,S.Z.著、隠れマルコフモデル及びレベルビルディングに基づく話者独立、シンタクス重視の結合ワード認識システム、IASSP,Vol.33,No.3,June 1985)。

特に認識エラーの確率を最小にするために、音声認識問題は次のように公式化できる。音響証拠観察(acoustic evidence observation)Ｏでは、音声認識の操作は、$W' = arg\ max P(W|O)$　　（１）

となるような最もありそうなワード列Ｗ'を見つけることである。ここで、最大化(maximization)による最大値はすべての可能なワード列Ｗ以上である。ベイズの

規則によれば、上記の方程式における事後確率P（W｜O）は以下のように書き換えられる。

$$P(W|O)=\frac{P(W)\,P(O|W)}{P(O)} \qquad (2)$$

ここで、P（O）は認識と無関係なので、ワード列の推定は以下の式で書ける。

$$W'=arg\ max\,P(W)\,P(O|W) \qquad (3)$$

ここで、P（W）はワード列Wが発音される事前確率を表し、P（O｜W）は、話者がワードシーケンスWを発音したときに、音響証拠Oが観察される確率である。P（O｜W）は音響パターンマッチングによって決定され、事前確率P（W）は使用される言語モデルによって定義される。

結合されたワード認識において、語彙が小さい（１００以下）ときは、言語におけるリーガルセンテンスを形成するために、どのワードが他のワードに論理的に続いているのかを厳密に把握するために確定的文法が使用される。確定的文法は可能性のあるワードの探索空間を暗に拘束して計算を大幅に減らすために、音響マッチングアルゴリズムに組み込むことが可能である。しかしながら、語彙のサイズが中ぐらい（１００より大、かつ、１０００より小さい）、あるいは、大きい（１０００よりも大）場合、ワードシーケンスW＝（$w_1, w_2, \cdots, w_n$）の確率は、確率的言語モデリングによって得られる。単純な確率理論により、事前確率P（W）は、

$$P(W)=P(w_1,w_2,\ldots,w_n)=\prod_{i=1}^{n}P(w_i|w_1,w_2,\ldots,w_{i-1}) \qquad (4)$$

のように分解できる。ここで、P（w_i｜$w_1, w_2, \cdots, w_{i-1}$）は、ワードシーケンス（$w_1, w_2, \cdots, w_{i-1}$）が話された後で$w_i$が話されたときの確率である。$w_i$の選択は入力ワードの全体の過去の履歴に依存する。語彙のサイズがVのとき、P（w_i｜$w_1, w_2, \cdots, w_{i-1}$）を完全に特定するために$V^i$値が必要となる。このことは、語彙のサイズが中ぐらいであっても、言語モデルをトレーニングするため

に、莫大な数のサンプルを必要とする。トレーニングが不十分なことによるP（$w_i \mid w_1, w_2, \cdots, w_{i-1}$）の不正確な推定は元の音響マッチングの結果を低下させてしまう。

上記の問題に対する実際的な解決は、w_iが（k－1）の先行するワード、$w_{i-1}, w_{i-2}, \cdots, w_{i-k+1}$のみに依存すると仮定することである。確率的言語モデルはk－グラム言語モデルが引き出されるP（$w_i \mid w_1, w_2, \cdots, w_{i-k+1}$）の条件で完全に記載することができる。k＞3ならば、たいていのワード列は言語内で発生しないので、ユニグラム（k＝1）、バイグラム（k＝2）、トリグラム（k＝3）が、文法を統計的に考慮する最も有効な確率的言語モデルである。言語モデリングはシンタクス(syntactic)及び意味(semantic)情報を含み認識上重要である。しかしながら、これらの確率は音声データの大規模な集積からトレーニングしなければならない。k－グラムがデータ内で発生しない場合など、利用可能なトレーニングデータが比較的制限されている場合は、P（$w_i \mid w_{i-2}, w_{i-1}$）はバイグラム確率P（$w_i \mid w_{i-1}$）から直接推定することができる。この工程の詳細は、F,Jelink著、実験的離散ディクテーション認識装置の開発、Proc.IEEE,vol.73,pp.1616-1624,1085)に開示されている。結合されたワード認識では、すべてのワードモデルが基本的な音声ユニットとして用いられ、連続音声認識では、音素、音節、半音節が基本的な音声ユニットとして用いられる。ワード復号器は適宜変更される。

従来の音声認識システムは分離能力の制限と、（電力消費、メモリの利用度などの）応用システムの制限と、通信チャネル特性を考慮することなしに、音響プロセッサとワード復号器とを一体化している。このことは、これらの2つの要素が適宜分離された分散音声認識システムを発明することにつながる。

本発明の要約

本発明においては、（ｉ）フロントエンド音響プロセッサがＬＰＣまたはフィルタバンクに基づいており、（ｉｉ）ワード復号器における音響パターンマッチングが隠れマルコフモデル（ＨＭＭ）、ダイナミックタイムワーピング（ＤＴＷ）、

あるいはニューラルネットワーク（ＮＮ）に基づいており、（ｉｉｉ）結合あるいは、連続的ワード認識のために、言語モデルが確定的あるいは確率的文法に基づいている改善された分散音声認識システムである。本発明は特徴抽出とワード復号の２つの要素を適宜分離することによって、システムのパフォーマンスを改善した点で、従来の音声認識装置とは異なっている。以下の例に示すように、セプストラム係数などのＬＰＣに基づく特徴が通信チャネルを介して送信される場合は、ＬＰＣとＬＳＰとの間の変換は特徴シーケンスへのノイズの影響を低減するために使用される。

図面の簡単な説明

本発明の特徴、目的、利点は、添付の図面を参照して以下の詳細な説明によって明らかになる。

図１は従来の音声認識システムのブロック図であり、

図２はワイヤレス通信環境における本発明の実施形態のブロック図であり、

図３は本発明の一般的なブロック図であり、

図４は、本発明の変換要素及び逆変換要素の実施形態のブロック図であり、

図５はローカルワードプロセッサとリモートワード検出器とを具備する本発明の望ましい実施形態のブロック図である。

望ましい実施形態の詳細な説明

標準的な音声認識装置において、認識またはトレーニング時、ほとんどの計算上の複雑さは音声認識装置のワード復号サブシステムに集中する。分散システムアーキテクチャを備えた音声認識装置においては、ワード復号タスクを、計算上の負荷を適宜吸収できるサブシステムに任せることが望ましい。信号処理による量子化誤差及び／またはチャネル誘引誤差の影響を低減するために、音響プロセッサはできるだけ音声源の近くに設けることが望ましい。

本発明の実施形態は図２に示される。この実施形態では、実行環境は、ポータブルセルラ電話またはパーソナル通信装置４０と、セル基地局４２としての中央通信センタとを具備するワイヤレス通信システムである。この実施形態では分散されたＶＲシステムが用いられる。分散ＶＲにおいては、音響プロセッサまたは

特徴抽出要素２２がパーソナル通信装置４０に設けられるとともに、ワード復号器４８が中央通信センタに設けられる。分散されたＶＲの代わりに、ＶＲがポータブルセルラ電話内で単独で実行される場合は、中間サイズの語彙で、結合されたワード認識であっても、高い計算コストのために実行不可能となってしまう。一方、ＶＲが単に基地局に設けられている場合は、音声コーデック及びチャネル効果に関連した音声の劣化によって、正確度が大きく低下してしまう。明らかに、提案された分散システム設計には３つの利点がある。第１は、電話４０には配置されないワード復号ハードウエアによって、セルラ電話のコストの低減が図れることである。第２は、計算負荷の大きいワード復号動作をローカルで実行することによるポータブル電話４０の（図示せぬ）電池の消耗が少なくなることである。第３は、分散システムの柔軟性及び延長性に加えて、認識の正確さが改善されることである。

音声がマイクロホン２０に供給されて音声信号が電気信号に変換され、特徴抽出要素２２に供給される。マイクロホン２０からの信号はアナログまたはディジタルである。アナログの場合は、アナログからディジタルへの変換器（図示せぬ）がマイクロホン２０と特徴抽出要素２２との間に挿入される。音声信号は特徴抽出要素２２に供給される。特徴抽出要素２２は入力音声の言語解釈を復元するのに使用される入力音声の関連する特性を抽出する。音声を推定するのに用いられる１つの特性は、入力音声フレームの周波数特性である。これは入力音声フレームの線形予測符号化パラメータとしてしばしば提供される。音声の抽出された特徴はトランスミッタ２４に供給して抽出特徴信号を符号化、変調、増幅した後、送受切り換え器２６を介してアンテナ２８に供給され、音声の特徴がセルラ基地局または中央通信センタ４２に送信される。既知の種々のディジタル符号化、変調、送信方法が用いられる。

中央通信センタ４２では、送信された特徴がアンテナ４４で受信されてレシーバ４６に供給される。レシーバ４６は受信された特徴に対して復調、復号を施してワード復号器４８に供給する。ワード復号器４８は音声の特徴から、音声の言語推定を決定してトランスミッタ５０にアクション信号を供給する。トランスミ

ッタ５０はこのアクション信号に対して増幅、変調、符号化を施して増幅された信号をアンテナ５２に供給する。アンテナ５２は推定されたワードまたはコマンド信号をポータブル電話４０に送信する。トランスミッタ５０は既知のディジタル符号化、変調、送信テクニックを実行する。

ポータブル電話４０では、推定されたワードまたはコマンド信号はアンテナ２８で受信される。アンテナ２８は受信信号を送受切り換え器２６を介してレシーバ３０に供給し、レシーバ３０はこの信号を復調、復号した後、コマンド信号または推定ワードを制御要素３８に供給する。受信コマンド信号または推定ワードに応答して、制御要素３８は意図する応答（例えば、電話番号をダイヤルする、ポータブル電話の表示スクリーンに情報を提供するなど）を提供する。

図２に示す同様のシステムは、中央通信センタ４２からの情報が送信された音声の解釈である必要はなく、中央通信センタ４２からの情報はポータブル電話によって送信された復号メッセージに対する応答である。中央通信センタ４２に通信ネットワークを介して結合された（図示せぬ）リモート応答システムに関するメッセージについて尋ねるときがあるが、この場合、中央通信センタ４２からポータブル電話４０へ送信された信号は、この実行においては応答マシンからのメッセージである。

特徴抽出要素２２を、中央通信センタ４２ではなくポータブル電話４０に設ける重要性は次の通りである。音響プロセッサが、分散ＶＲに対向して、中央通信センタ４２に設けられたとき、低帯域ディジタル無線チャネルは、量子化ひずみによる特徴ベクトルの解像度を制限する（第１のサブシステムにおける）ボコーダを必要とする。しかしながら、音響プロセッサをポータブルまたはセルラ電話に設けることによって、すべてのチャネル帯域を特徴の送信のために使うことができる。概して、抽出された音響特徴ベクトルは送信のために音声よりも帯域を必要としない。認識の正確度は入力音声信号の劣化に大きく依存するので、特徴抽出要素２２をできるだけユーザに近接させる必要があり、これによって、特徴抽出要素２２は、送信中にさらに破壊されるボコーダによって処理された(vocoded)電話音声の代わりにマイクロホン音声に基づいて特徴ベクトルを抽出する。

実際上は、音声認識装置は背景雑音などの周囲の条件下で動作するように設計される。すなわち、雑音の存在下での音声認識の問題を考慮することが重要である。語彙（基準パターン）のトレーニングがテスト時の条件と全く（またはほぼ）同じ環境で実行されれば、音声認識装置は雑音が多い環境においてもよいパフォーマンスが得られるとともに、雑音によって認識の正確度が大きく劣化するのを低減することができる。トレーニングとテスト条件との間の不整合は認識のパフォーマンスにおける主な劣化要因の１つである。（前記したように音響特徴の方が音声信号よりも送信時の帯域を必要としないので）、音響特徴が音声信号よりもより大きな信頼度で通信チャネルを横断できると仮定すると、提案された分散音声認識システムは整合された状態を提供するのにより適している。音声認識装置がリモート状態で実行されたとき、ワイヤレス通信において発生する主にフェージングなどのチャネルバリエーションのために、整合状態が大きく破壊される。大規模なトレーニング計算がローカルで吸収されるなら、ＶＲをローカルで実行することによりこれらの影響を避けることができる。不幸なことに、多くの応用ではこれは不可能である。明らかに、分散音声認識の構成はチャネルの複雑さによって起こる不整合の状態を避けて、中央集権構成の欠点を補うことができる。

図３において、ディジタル音声サンプルは特徴抽出要素５１に供給される。特徴抽出要素５１は通信チャネル５６を介して特徴をワード推定要素６２に供給し、ここで推定ワード列が決定される。音声信号は各音声フレームに対する特徴を決定する音響プロセッサ５２に供給される。ワード復号器は認識とトレーニングの作業に対する入力として音響特徴シーケンスを必要とするので、これらの特徴は通信チャネル５６を介して送信される必要がある。しかしながら、通常の音声認識システムにおいて用いられる特徴が雑音の多いチャネルを介した送信に適しているわけではない。例えば、変換要素２２は音声源符号化(source encoding)を行ってチャネル雑音の影響を低減する必要がある。音声認識装置で広範に用いられているＬＰＣに基づく音響特徴の１つはケプストラム係数、$\{c_i\}$である。これはＬＰＣ係数、$\{a_i\}$から直接次のようにして得ることができる。

$$c_m=a_m+\sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k} \qquad m=1,\ldots,P \qquad (5)$$

$$c_m=\sum_{k=1}^{m-1}\left(\frac{k}{m}\right)c_k a_{m-k} \qquad m=P+1,\ldots,Q \qquad (6)$$

ここで、Ｐは使用されるＬＰＣフィルタの次数であり、Ｑはケプストラム特徴ベクトルのサイズである。ケプストラム特徴ベクトルは急峻に変化するので、ケプストラム係数のフレームシーケンスを圧縮することは容易ではない。しかしながら、ＬＰＣと、ゆるやかに変化し、デルタパルス符号変調（ＤＰＣＭ）によって効率的に符号化できる線スペクトルペア（ＬＳＰ）周波数との間の変換が存在する。ケプストラム係数はＬＰＣ係数から直接引き出すことができるので、ＬＰＣは変換要素５４によってＬＳＰに変換され、ここで通信チャネル５６を横断すべく符号化される。リモートワード推定要素６２では、変換された特徴が逆変換要素６０によって逆変換されて音響特徴がワードプロセッサ６４に供給され、ワードプロセッサ６４はこれに応答して推定ワード列を提供する。

変換要素５４の実施形態は図４に変換サブシステム７０として示されている。図４において、音響プロセッサ５２からのＬＰＣ係数は、ＬＰＣからＬＳＰ変換要素７２に供給される。ＬＰＣからＬＳＰ変換要素７２において、ＬＳＰ係数は次の通りに決定される。Ｐ次の次数のＬＰＣ係数に対して、対応するＬＳＰ周波数が次の方程式の０とπの間に存在するＰ個の根として得られる。

$$P(w)=\cos 5w+p_1\cos 4w+\cdots+p_5/2 \qquad (7)$$

$$Q(w)=\cos 5w+q_1\cos 4w+\cdots+q_5/2 \qquad (8)$$

ここで、p_iとq_iは帰納的に次のように求められる。

$$p_0=q_0=1 \qquad (9)$$

$$p_i=-a_i-a_{p-i}-p_{i-1},\ 1\leqq i\leqq P/2 \qquad (10)$$

$$q_i=-a_i+a_{p-i}-q_{i-1},\ 1\leqq i\leqq P/2 \qquad (11)$$

ＬＳＰ周波数はＤＰＣＭ要素７４に供給されて通信チャネル７６を介しての送信のために符号化される。

逆変換要素７８において、チャネルからの受信信号は、音声信号のＬＳＰ周波数を復元すべく、逆ＤＰＣＭ要素８０とＬＰＣからＬＳＰ要素８２とを通過される。ＬＰＣからＬＳＰ要素７２の逆プロセスは、ＬＳＰ周波数をケプストラム係数を引き出すのに用いられるＬＰＣ係数に変換するＬＳＰからＬＰＣ要素８２によって実行される。ＬＳＰからＬＰＣ要素８２は次のように変換を実行する。

$$P(z) = (1+z^{-1})\prod_{i=1}^{P/2}(1-2\cos(w_{2i-1})z^{-1}+z^{-2}) \quad (12)$$

$$Q(z) = (1-z^{-1})\prod_{i=1}^{P/2}(1-2\cos(w_{2i})z^{-1}+z^{-2}) \quad (13)$$

$$A(z) = 1-\sum_{i=1}^{P}a_i z^{-i} = \frac{P(z)+Q(z)}{2} \quad (14)$$

ＬＰＣ係数はＬＰＣからケプストラム要素８４に供給され、ここで、方程式５及び６に応じてケプストラム係数をワード復号器６４に供給する。

ワード復号器は、通信チャネルを介して直接送信されたときに雑音の影響を受けやすい音響特徴シーケンスのみに依存するので、音響特徴シーケンスが引き出されて図３に示すようなサブシステム５１において送信を可能にする代替表現に変換される。ワード復号器で使用される音響特徴シーケンスは後で逆変換によって得られる。すなわち、ＶＲの分散構成においては、空中（チャネル）を介して送信された特徴シーケンスはワード復号器において実際に使用されるものとは異なっている。変換要素７０からの出力は既知の種々のエラー保護方法によってさらに符号化される。

本発明の改善された実施形態が図５に示されている。ワイヤレス通信への応用においては、ユーザは、部分的に高価なチャネルアクセスのために、小数の単純だが供給に用いられる音声コマンドに対する通信チャネルを占有しないことを望む。これは、比較的小さい語彙サイズをもつ音声認識装置がローカルで送受話器において実行されるとともに、大きな語彙サイズをもつ第２の音声認識システムがリモート基地局に設けられるという点を考慮すると、送受話器１００と基地局１１０との間にワード復号機能を分散させることによって達成される。それらは

送受話器において同じ音響プロセッサを共有する。ローカルのワード復号器の語彙テーブルは最もよく用いられるワード、またはワード列を含む。一方、リモートのワード復号器の語彙テーブルは正規のワード、またはワード列を含む。このような構成に基づいて、図５に示すように、チャネルがビジーである平均時間を小さくして認識の正確度を増大させることができる。

さらに、２群の音声コマンドが利用され、第１は特殊音声コマンドと呼ばれ、ローカルＶＲによって認識できるコマンドに対応する。第２は正規の音声コマンドと呼ばれ、ローカルＶＲによって認識されないコマンドに対応する。特殊な音声コマンドが発音されるときはいつでも、真の音響特徴がローカルワード復号器のために抽出され、音声認識機能は通信チャネルにアクセスすることなしにローカルで実行される。正規の音声コマンドが発音されるとき、変換された音響特徴ベクトルがチャネルを介して送信され、復号化が基地局においてリモートで行われる。

特殊な音声コマンドに対する音響特徴は変換、あるいは符号化される必要がなく、ローカルのＶＲに対する語彙サイズは小さいので、要求される計算量はリモートのものよりもはるかに小さい（語彙の中から正確なワード列を探索するときの計算量は語彙のサイズに比例する）。さらに、音響特徴はチャネル内での破壊なしにローカルＶＲに直接供給されるので、ローカルの音声認識装置はリモートＶＲに比較して（状態数が小さい、状態出力確率などに対する混合要素の数が小さいなど）ＨＭＭの単純化された形態によって構成される。これは制限された語彙で送受信機（サブシステム１）でのＶＲのローカル構成を可能にし、この場合の計算量は制限されたものとなる。分散されたＶＲ構成はワイヤレス通信システム以外の他の応用分野にも適用可能である。

図５において、音声信号は音響プロセッサ１０２に供給されて、音声信号から例えばＬＰＣに基づく特徴パラメータなどの特徴が抽出される。これらの特徴はローカルのワード復号器１０６に供給されて、入力音声信号を小さな語彙から識別するための探索が行われる。ワード復号器１０６が入力ワード列を復号できず、リモートのＶＲが復号すべきであるときは、特徴を送信する準備をする変換要

素１０４に信号を送る。変換された特徴は通信チャネル１０８を介してリモートのワード復号器１１０に送信される。変換された特徴は逆変換要素１１２に供給される。この逆変換要素１１２は変換要素１０４の逆変換を実行してリモートのワード復号器要素１１４に音響特徴を供給する。ワード復号器要素１１４はこれに応答して推定リモートワード列を提供する。

好ましい実施形態の前記した説明は当業者が本発明を製造または使用可能なように提供される。上記の実施形態に対する種々の変形が可能であり、ここに定義された一般的原理は発明に相当する能力を用いることなしに他の実施形態に適用可能である。すなわち、本発明は上記の実施形態に制限されることはなく、ここに開示された原理と新規な特徴に一致する範囲で広範な権利範囲が与えられるべきである。

【図１】

従来の技術

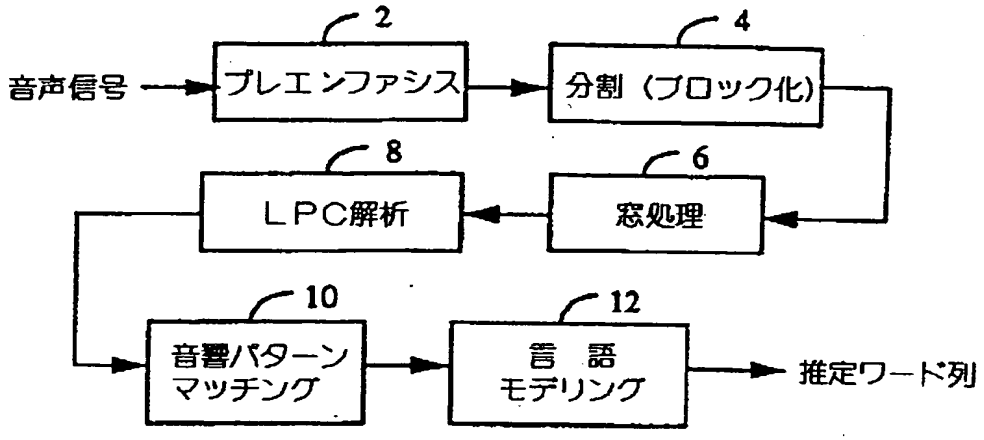

FIG. 1

【図３】

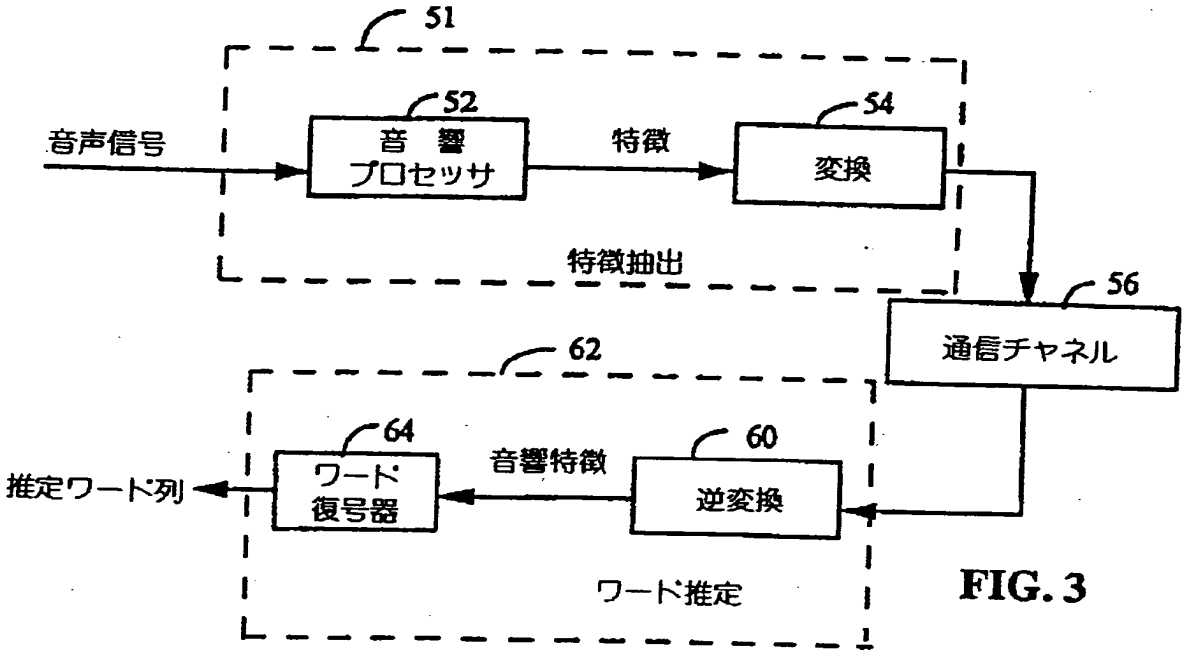
51
52
音声信号
音響プロセッサ
特徴
54
変換
特徴抽出
56
通信チャネル
62
64
推定ワード列
ワード復号器
音響特徴
60
逆変換
ワード推定

FIG. 3

【図２】

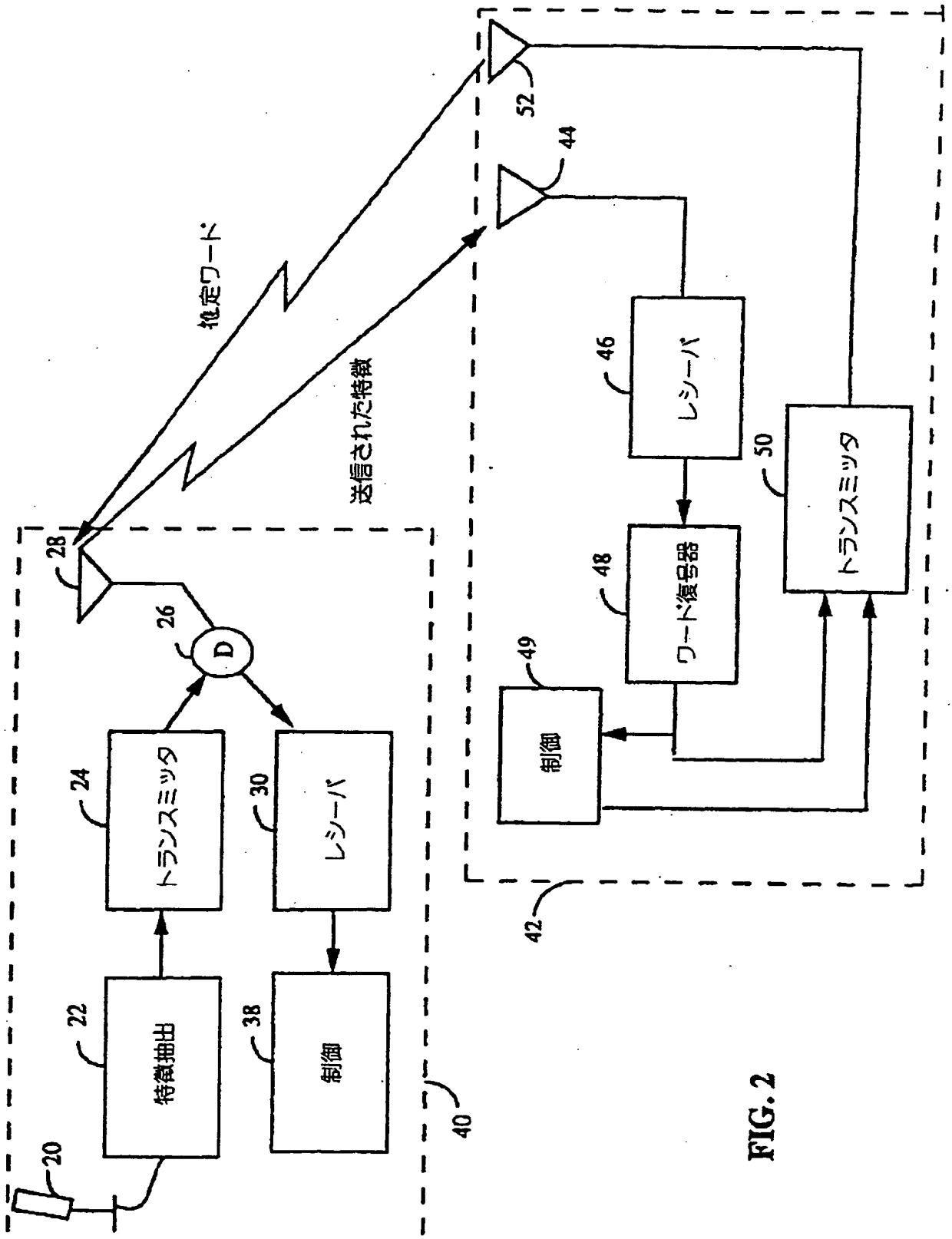
推定ワード
送信された特徴
28
26
D
24
トランスミッタ
30
レシーバ
22
特徴抽出
38
制御
20
40
52
44
46
レシーバ
50
トランスミッタ
48
ワード復号器
49
制御
42

FIG. 2

【図４】

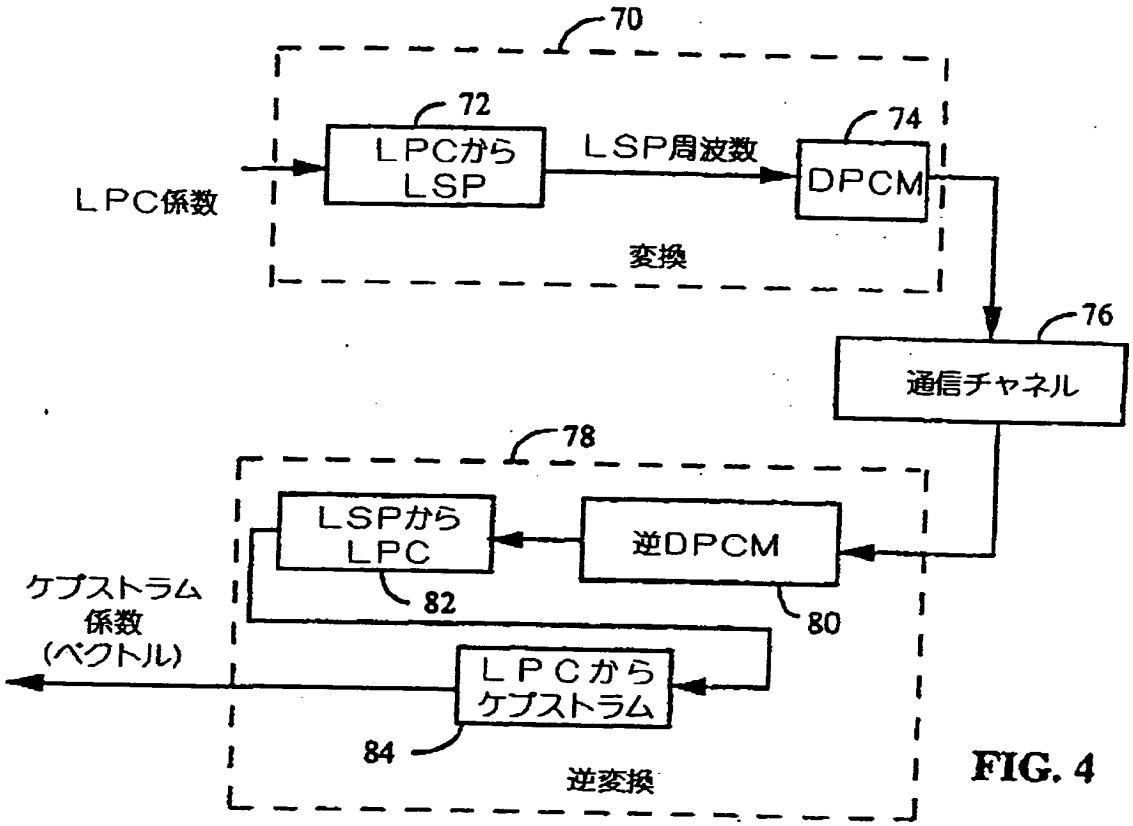

FIG. 4

【図５】

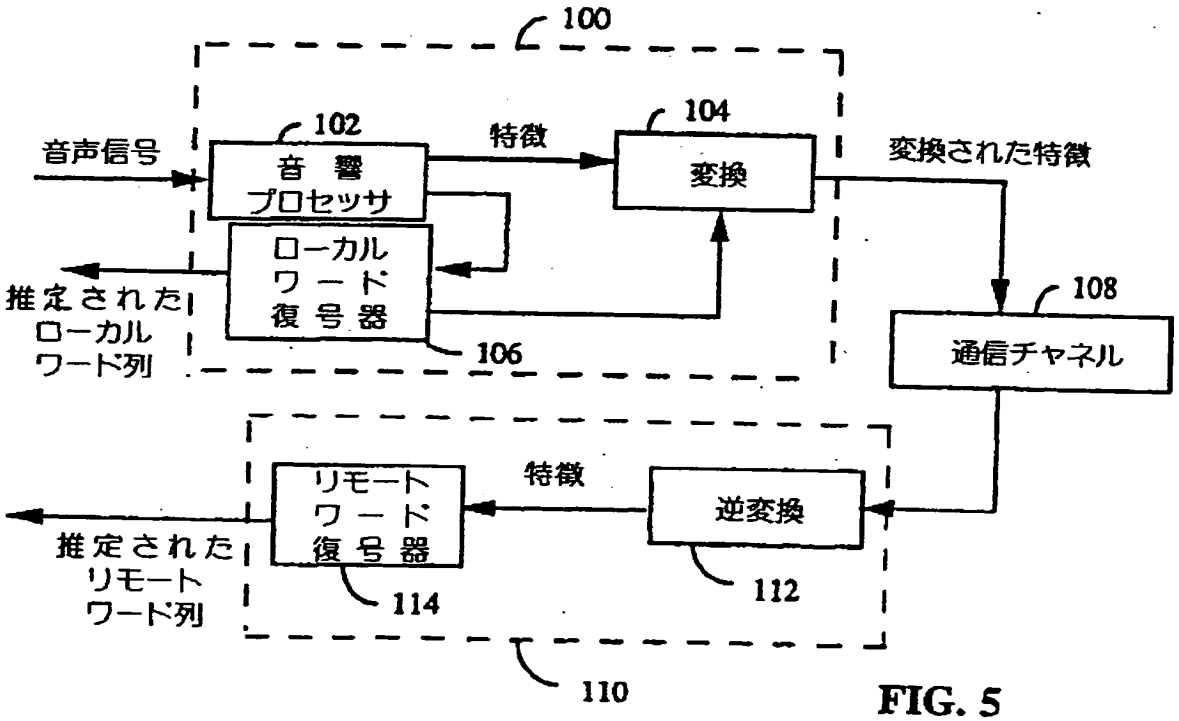

FIG. 5

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 94/14803

A. CLASSIFICATION OF SUBJECT MATTER

G 10 L 5/06,G 10 L 5/02,G 10 L 5/00,G 10 L 7/08,
G 10 L 9/06

According to International Patent Classification (IPC) or to both national classification and IPC6

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G 10 L,G 06 F,H 04 M

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP, A, 0 534 410 (NIPPON HASO KYOKAI) 31 March 1993 (31.03.93), fig. 1; abstract; claim 1. --	1
A	EP, A, 0 177 405 (REGIE NATIONALE DES USINES RENAULT) 09 April 1986 (09.04.86), fig. 1; abstract; claim 1. --	1
A	EP, A , 0 108 354 (INTERNATIONAL STANDARD ELECTRIC CORPORATION) 16 May 1984 (16.05.84), fig. 1; abstract; claim 1. ----	1

☐ Further documents are listed in the continuation of box C.　　☐ Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 March 1995	07. 04. 95
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Authorized officer BERGER e.h.

Form PCT/ISA/210 (second sheet) (July 1992)

フロントページの続き

(81)指定国　　EP(AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OA(BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), AP(KE, MW, SD, SZ), AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, JP, KE, KG, KP, KR, KZ, LK, LR, LT, LU, LV, MD, MG, MN, MW, NL, NO, NZ, PL, PT, RO, RU, SD, SE, SI, SK, TJ, TT, UA, UZ, VN